



Assessment of Orofacial Strength in Patients with Dysarthria

Authors:

Nancy Pearl Solomon, **Heather M. Clark**, Matthew J. Makashay and Lisa A. Newman

No Abstract

Solomon, Nancy, **Clark, Heather**, Makashay, Makashay, & Newman, Lisa. (2008). Assessment of orofacial strength in adults with dysarthria. *Journal of Medical Speech-Language Pathology*, (16(4)) (Dec 2008): 251-258. Published by Delmar Learning (ISSN: 1065-1438).

Solomon, N.P., Clark, H. M., Makashay, M.J., & Newman, L.A. (2008). Assessment of orofacial strength in adults with dysarthria. **Journal of Medical Speech-Language Pathology**, 16(4) (Dec 2008): 251-258. Published by Delmar Learning (ISSN: 1065-1438).

Assessment of Orofacial Strength in Patients with Dysarthria

Nancy Pearl Solomon, Heather M. Clark, Matthew J. Makashay and Lisa A. Newman

Assessment of orofacial weakness is common during the evaluation of patients with suspected dysarthria. This study addressed the validity of clinical assessments of orofacial weakness by comparing clinical (subjective) ratings to instrumental (objective) measures. Forty-four adults referred to a speech pathology clinic for dysarthria evaluation were tested for strength of the tongue during elevation, lateralization, and protrusion, and for the strength of the muscles of the lower face during buccodental and interlabial compression. Subjective assessment of weakness involved rating maximum resistance against a firmly held tongue depressor, using a 5-point scale. Objective assessment involved the Iowa Oral Performance Instrument (IOPI), measured as the maximal pressure generated against an air-filled bulb. A recent adaptation to the IOPI permitted testing of tongue and cheek strength using tasks that are comparable to the subjective tasks. Moderate correlations were found between the objective and subjective evaluations, with the strongest correlations for tongue lateralization. Lower pressure values were associated with higher subjective ratings of weakness for each task, although there was substantial overlap in the data. These results, combined with the notion that examiner bias is inherent to clinical assessment, support the use of instrumentation to improve objectivity and precision of measurement in the clinic.

The muscles of the tongue and face are integral to the functions of speech, facial expression, eating, and swallowing. Therefore, speech-language pathologists (SLPs) assess orofacial function as part of a diagnostic evaluation for patients with suspected dysarthria or dysphagia. Maximum performance tasks, such as those used for strength testing, are often included in such evaluations, because they may reveal a neuromuscular impairment, serve as a diagnostic aid, and provide information to facilitate treatment planning. The specific relationship between orofacial weakness and functional activities is not known and cannot be fully evaluated without a reliable, accurate, and valid measurement technique.

Typically, orofacial muscular strength is assessed by the SLP asking the patient to push the tongue or lips against a resistance provided by a tongue depressor or a finger. Strength is then rated as normal or according to a scale for weakness, such as the ordinal categories of mild, moderate, and severe. There are no norms for this test, and ratings are based on the clinician's experience and internal representation of strength and weakness. Rarely are instrumental measures used in the clinic, although they are available (Bu Sha et al., 2000; Hayashi et al., 2002; Robin et al., 1991; Thompson et al., 1995).

A commercially available instrument for this purpose is the Iowa Oral Performance Instrument (IOPI Northwest). The IOPI has been available to study tongue strength since the early 1990s and has been used in multiple research studies and selected clinics. Standard testing involves placing an oblong, air-filled, soft plastic bulb along the hard palate just behind the central incisors. The patient elevates the anterior tongue dorsum to press against the bulb as hard as possible. Compressing the air within the bulb increases pressure, which is sensed by the IOPI's pressure-transducing circuitry. The peak pressure generated indicates strength.

Clark et al. (2003) tested 63 patients with a variety of medical diagnoses referred for evaluation of possible dysphagia using the IOPI's tongue-elevation task, although the positioning of the tongue bulb may have underestimated the actual output. These results were compared to subjective ratings of weakness by experienced clinicians and student clinicians, who had the participants protrude and lateralize the tongue against resistance provided by a tongue depressor. The examiners determined a single rating of normal, slightly weak, moderately weak, or severely weak. Comparisons of the two types of evaluation revealed a moderate correlation (Spearman's correlation, $[r_{\text{sub.s}}] = .541$) between the subjective and objective data. Interestingly, correlations were stronger for the student clinicians ($[r_{\text{sub.s}}] = .696$) than for the experienced clinicians ($[r_{\text{sub.s}}] = .395$). The authors speculated that experienced SLPs may have given greater emphasis to tongue lateralization when rating strength, a task not assessed by the IOPI in their study. One might expect to find higher correlations between subjective and objective measures of tongue strength if the tasks used to make each assessment were comparable.

Wood et al. (1992) assessed medial lip compression over 5 days in healthy adults and persons with Parkinson's disease, multiple sclerosis, or stroke. The instrumentation involved a strain-gauged cantilever beam positioned between the lips at midline. The clinical assessment was based on squeezing a tongue depressor between the lips while the examiner attempted to remove it; resistance was rated along a 4-point scale. The correlation coefficients between

these relatively comparable assessments of interlabial compression were .67 for the upper lip and .62 for the lower lip. These stronger correlations could be attributed to the comparability of the tasks used for the two assessments. They also found that the objective assessment was quite reliable and did not reveal any particular learning or fatigue effects.

Recent innovations in orofacial strength assessment permit measures of tongue strength during tasks other than anterior tongue-dorsum elevation. An accessory to the IOPI is a holder designed to hold the tongue bulb on its side (vertically). This allows for testing of tongue lateralization and protrusion (Clark, Obrien, Newcomb, & Calleja, submitted; Luschei, 2008). Furthermore, the holder can be used to position the bulb within the buccal cavity, allowing for assessment of cheek compression, presumably as a result of buccinator and risorius muscle contraction. A final adaptation of IOPI testing was to sandwich a tongue bulb between two tongue depressors to assess medical lip compression. Given these simple adaptations, the present study was possible. Its purpose was to compare objective measures to subjective ratings taken during comparable tasks for orofacial-strength testing in persons with dysarthria.

METHOD

Participants

Potential candidates were referred to the Speech Pathology Clinic at a military treatment facility for speech or swallowing evaluations. Of 51 patients so identified and referred to the principal investigator, 44 had dysarthria and were included as participants. The remaining candidates were excluded because their disorders had resolved or were not neurogenic. Participants were 40 men and 4 women who ranged in age from 18 to 78 years ($M = 39.6$, $SD = 16.7$). They presented with a variety of medical diagnoses, including neurovascular event or neuroplasm ($n = 18$), head/neck/brain injury ($n = 12$), three of whom also had strokes), progressive and/or generalized neurologic diseases ($n = 11$), and other ($n = 3$; infectious, autoimmune, endocrine). All but one of the head/neck/brain-injured patients had gunshot or shrapnel wounds and/or were exposed to blast explosions in a combat environment. All candidates provided informed consent according to the institution's human use committee.

Procedures

Orofacial strength was assessed objectively and subjectively. Tasks for both assessments included:

1. anterior tongue-dorsum elevation.
2. posterior tongue-dorsum elevation.
3. tongue protrusion,
4. tongue right lateralization,

5. tongue left lateralization,
6. interlabial compression,
7. right cheek compression, and
8. left cheek compression.

The order of the types of assessments was alternated among participants, and the order of the tasks within each assessments was randomized.

Objective Assessments

Strength was tested with the IOPI and its accessories (tongue bulb, tongue-bulb holder). Participants pushed as hard as possible against the air-filled IOPI bulb. Anterior tongue-dorsum elevation was tested according to standard procedures--the bulb was placed along the hard palate immediately behind the central incisors. The bulb was placed more posteriorly along the hard palate for posterior tongue-dorsum elevation. For tongue protrusion, the tongue bulb was adhered to the holder with double-sided surgical-grade tape so that the bulb faced the tongue in the anterior oral cavity. The holder was held in place by placing its cushioned bite surface, made of polyoxymethylene (Delrin [R] by DuPont), between the central incisors. The holder was held in place by the molars so that the bulb was in the lateral oral cavity for tongue lateralization, or turned so that the bulb was in the buccal cavity for cheek compression. For more information about the lateral tongue-bulb holder, refer to Clark et al. (submitted) and Luschei (2008). For tongue protrusion, lateralization, and posterior elevation, the tongue tended to slip, so the IOPI bulb and its adaptor were wrapped in a single layer of sterile gauze. Interlabial compression was assessed by sandwiching an IOPI bulb between two tongue depressors and placing the tips of the depressors between the lips at midline; participants gently placed their teeth together to avoid assistance by the jaw-closing muscles before squeezing their lips together. Each task was performed three times and the maximum pressure generated was recorded.

Subjective Assessments

A single examiner, an SLP for 23 years, conducted the subjective assessments. She was blinded to the objective results when the objective assessment was conducted first. Participants were instructed to push as hard as possible against a tongue depressor while the examiner provided resistance. Care was taken to place the tongue depressor in a position similar to the IOPI bulb's position for the objective tasks. For tongue protrusion, the tongue depressor was oriented vertically just anterior to the teeth. Participants were instructed to protrude the tongue against the depressor as hard as possible while the examiner held the free end tightly. For anterior and posterior tongue elevation, the tongue depressor was oriented horizontally and held firmly on top of the anterior third of the tongue and the middle of the tongue, respectively, while the participant attempted lingual elevation. The depressor was oriented vertically within the

mouth and placed next to the molars while the participant pushed the tongue laterally against resistance. For cheek strength, the examiner displaced the cheek laterally with the tongue depressor and had the participant pull the cheek in towards the teeth. Lip strength was assessed by placing the depressor horizontally between the lips (and without involving the teeth), asking the participant to press the lips tightly against the depressor while the examiner attempted to displace the stick superiorly and inferiorly. Ratings for the upper lip and lower lip were averaged for analysis. Tasks were repeated 2-3 times, and the examiner rated the best attempt along a 5-point scale (1 = normal, 2 = mild, 3 = moderate, 4 = severe, 5 = profound weakness).

Reliability

Fifteen participants returned for repeated objective assessment and 13 received repeated subjective assessment with the original examiner on a separate day. A second examiner, an SLP for over 30 years, conducted subjective assessments on 11 participants for interrater reliability. It should be noted that these reliability measures are affected not only by examiner variability but also by performance variability by the participant.

RESULTS

Moderate correlations between objective and subjective assessments of orofacial strength were found across tasks ($[r_{\text{sub}.3}] = -.449$ to $-.719$; Table 1), with relatively weaker associations for the facial muscles than for the tongue, in general. The strongest associations were for tongue lateralization. These relationships can be seen in Figure 1, which illustrates individual data for each task. Objective measures of strength (the maximum pressure generated across 3 trials, in kPa) are plotted against the subjective ratings for each task. There is obvious overlap of the objective data across the subjective ratings, although the plots support the finding of decreased maximum pressure as subjective rating increases, especially when weakness was judged to be at least moderate in severity. It should be noted that participants are not evenly distributed across the severity scale, a factor that could weaken correlations.

TABLE 1. Spearman's correlations (r_s) between objective and subjective assessments for 44 adults with dysarthria.

	r_s	p
Tongue		
Elevation		
Anterior	-.589	< .001**
Posterior	-.449	.008**
Lateralization		
Right	-.643	< .001**
Left	-.719	< .001**
Protrusion	-.605	< .001**
Lip compression	-.465	.002**
Cheek compression		
Right	-.555	< .001**
Left	-.466	.002**

** $p < .01$

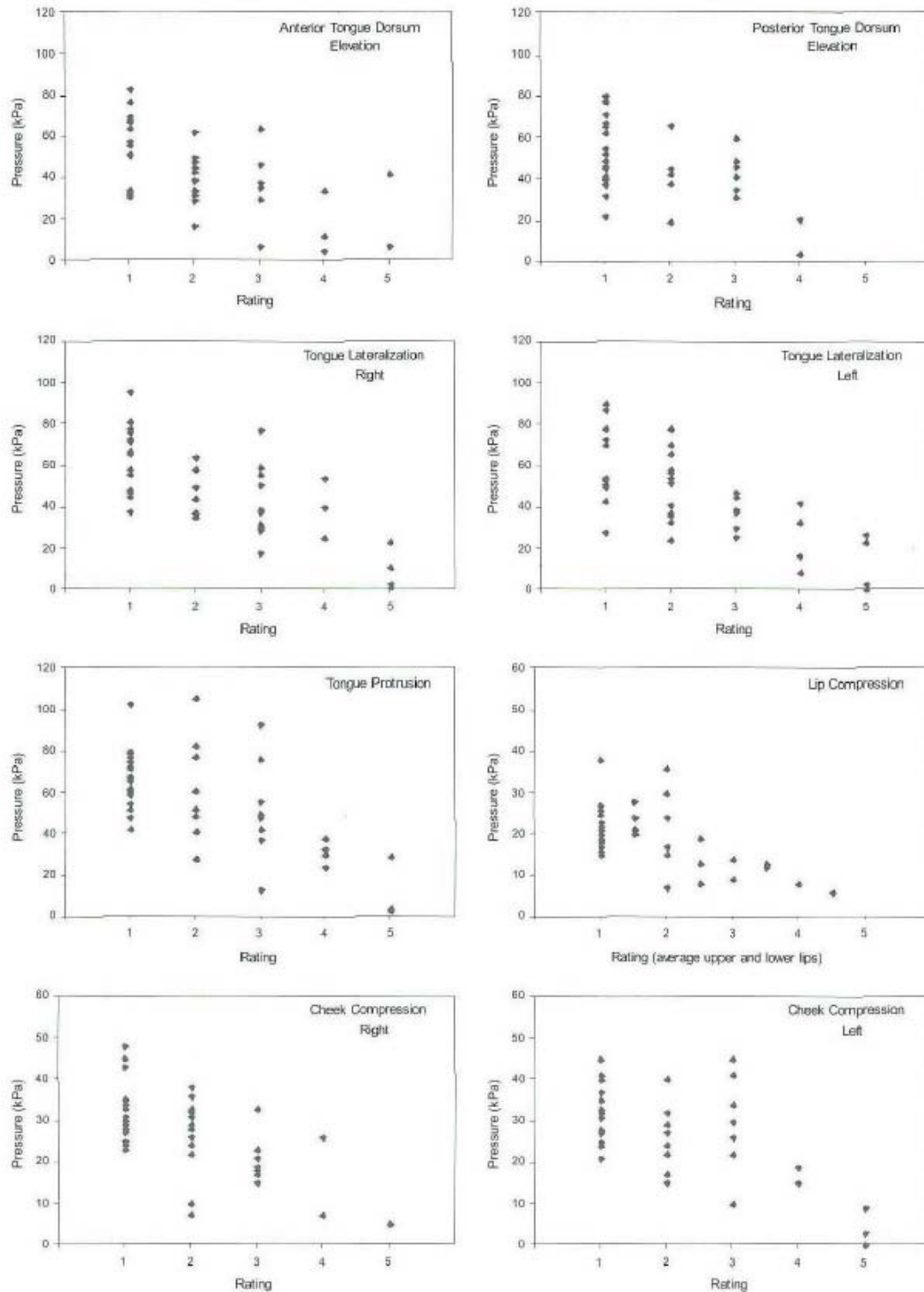


Figure 1. Objective data (maximum pressure, in kPa, obtained across three trials) plotted against subjective ratings (1–5, where 1 = normal and 5 = profound weakness). The individual plots represent strength assessments (maximum effort contractions) for each of the eight tasks. Note that the scale for pressure differs between the tasks for the tongue and those for the facial muscles.

Reliability analysis based on a subset of participants indicated that objective measures of strength are greater during a second assessment as compared to the first [RM-ANOVA, $F(1,12) = 6.83$, $p = .023$] (Table 2). The average difference in maximum pressure pooled for all tasks was 2.9 kPa. As listed in Table 2, pressures increased from the first to the second assessment from 0.7 kPa (for the lips) to 5.9 kPa (for anterior tongue elevation), on average; although the aggregate analysis indicated a statistically significant difference, none of the individual tasks did.

TABLE 2. Reliability of objective measures (pressure, in kPa) across two testing visits. Mean (and SD) difference for Visit 2–Visit 1 across 15 participants; inferential statistical analysis by RM-ANOVA.

	Mean Difference	(SD)	<i>F</i>	<i>p</i>
Overall	2.9	(8.2)	6.83	.023*
Tongue				
Elevation				
Anterior	5.9	(11.4)	3.95	.067
Posterior	2.7	(8.9)	1.34	.266
Lateralization				
Right	1.8	(11.1)	0.33	.577
Left	3.2	(9.8)	1.34	.269
Protrusion	1.9	(7.8)	0.86	.369
Lips	0.7	(3.8)	0.56	.466
Cheek				
Right	4.1	(6.7)	5.32	.038
Left	2.7	(6.0)	2.87	.114

* $p < .05$ for overall analysis; $p < .006$ with Bonferroni correction for multiple tasks.

Reliability analysis of subjective ratings across two assessments is listed in Table 3. With the tasks pooled, intraclass correlation coefficients (ICC) were .760 ($F = 7.56$, $p < .001$) for intrarater reliability, and .535 ($F = 3.56$, $p < .001$) for interrater reliability. Intrarater agreement for subjective assessments within 1 value on the 5-point scale was 67%, and within 2 scale values was 94%. Interrater agreement was 47% within 1 scale value, and 82% within 2 scale values.

TABLE 3. Intraclass correlation coefficients for reliability of subjective assessment across different testing sessions for 13 (intrarater) and 11 (inter-rater) participants.

	Intrarater	Interrater
Overall	.760	.535
Tongue		
Elevation		
Anterior	.750	.464
Posterior	.841	.539
Lateralization		
Right	.782	.484
Left	.673	.591
Protrusion	.709	.766
Lip		
Upper	.854	.575
Lower	.484	.492
Cheek		
Right	.609	-.012
Left	.855	.815

DISCUSSION

Validating clinical assessments of orofacial strength against a quantitative, objective measure is critical for accurate and meaningful clinical evaluation and management. In this study, objective and subjective assessments of orofacial strength were found to be moderately, but not strongly, correlated for 44 patients with dysarthria. The strength of the correlations was similar to those reported previously by Clark et al. (2003) for tongue strength, despite their use of tasks that differed between assessments. Wood et al. (1992) reported stronger correlations than those reported here for medial lip compression. Tasks only differed slightly, as both studies had participants hold a tongue depressor between maximally compressed lips. In the study by Wood et al., examiners rated resistance as they attempted to pull the depressor out; we attempted to displace the lips vertically with the tongue depressor. It is difficult to ascertain whether this procedural difference can explain the discrepancy in findings. Another, perhaps more important, difference is that Wood et al. rated lip strength clinically after taking objective measures. In the

present study, we alternated the order of objective and subjective assessments, and the examiner was blinded to the objective results when rating strength.

Examiner bias is a potential problem in any clinical assessment study. It is usually obvious that an impairment exists from looking at the patient or having a brief conversation. The experienced clinician will start developing hypotheses about the nature and severity of the dysarthria within the first few minutes of meeting the patient. This may affect the examiner's expectations regarding the presence of orofacial weakness. Another potential bias relates to the perceived abilities of the patient. If the patient appears frail, the clinician may avoid providing as much physical resistance to the orofacial structures as she or he might do if the patient appears younger and healthier. Using an objective instrument can help clinicians avoid such pitfalls.

One difference between previous studies and the present study is patient population. Because these data were collected at a military medical facility, many of the participants were active-duty soldiers. Other patients at this facility are family members or retired military officers. The active-duty patients are generally young, male, and highly physically fit. If their impairment is due to combat injury, it may involve multiple mechanisms and systems (e.g., shrapnel penetration, traumatic brain injury, stroke) rather than a single disease process. Patient populations involving neurogenic disorders in civilian medical facilities tend to be more gender balanced and to comprise older, less physically fit individuals. Because variations in results between facilities may be related to population effects, participant characteristics should be noted when comparing studies.

Test-retest reliability may be impacted by differences in patient populations. Clark et al. (2003) reported excellent agreement within and between experienced examiners when assessing tongue strength in participants who were generally older than in the present study. In our facility, we see many acute injuries in younger, premorbidly healthier individuals who tend to recover quickly from certain injuries. Our test-retest reliability was markedly lower, which could be due to patient-performance differences and/or examiner-interpretation differences. Careful examination of the present interrater reliability data revealed that the majority (83%) of the differences that were greater than 1 scale value were in the direction of more severe ratings by the original examiner. The second examiner claimed to provide less resistance than the original examiner, thereby not pushing the participant to his or her maximum potential. This followed from her philosophy that the purpose of the evaluation was to determine normal function rather than maximum performance. Clearly, Maximal strength far exceeds the requirements for normal speech or swallowing, although the extent to which weakness specifically impacts these functions remains to be determined and is a topic of future research.

Reliability of the subjective ratings may also be affected by the number of values on the scale. Both Clark et al. (2003) and Wood et al. (1992) used 4-point scales. Evident from the scatterplots provided here is that the 5th scale value was rarely used. Agreement would be enhanced and clinical interpretation might be improved with the use of the typical 4-point scale (normal, mild, moderate, and severe).

Performance variables degrade test-retest reliability of the objective data. The participant may not be willing to push as hard as possible, either because of a (subconscious) desire to appear impaired, being inhibited by pain or by the fear of inflicting pain, or not fully understanding instructions. The objective data for reliability revealed that greater pressures were generated on a second day of testing. No particular task differed significantly over the two days when examined individually, and the average differences were small. Previous literature on this issue is mixed. Barlow and Muller (1991), Wood et al. (1992), and Barlow and Rath (1985) found no learning or fatigue effects over 5 days of assessing lip strength. Clark et al. (2008) reported a trend, but not a statistically significant difference, for strength to increase across two sessions. Alternately, O'Day et al. (2005) reported a significant increase in tongue-elevation strength over 3, but not 4, days of testing. Although the day-to-day differences in orofacial strength testing results do not appear to be robust, it would be prudent to assess strength across 2 or 3 days whenever clinically practical.

Examination of the strength data from the IOPI reveals that the tongue generates a similar amount of pressure regardless of the direction of movement. Patients judged as having normal tongue strength tended to generate pressures that centered around 60 kPa. If substantiated through additional research, this may indicate that testing only a small number of tasks with the tongue may be adequate for assessing tongue strength. Of course, separate assessments for tongue lateralization would be necessary for patients with obvious unilateral impairments.

The overarching goal of this line of research is to determine the optimal method for assessing orofacial strength in a clinical environment. Although convenient, clinical ratings are problematic due to their subjectivity and imprecision. Clinician bias, experience, and task interpretation are inherent for such procedures, but using objective techniques can eliminate these factors. The clinician still holds the responsibility for providing clear instructions, proper placement of the measurement tool, external motivation, and accurate data recording, but it is the patient's responsibility to generate the best possible performance. Use of an instrument like the IOPI provides feedback, further motivating the patient. Once weakness is accurately assessed, it can be examined for its impact on function. Although normal strength and profound weakness have obvious effects on function, sensitive measurement tools may reveal a "threshold" of weakness beyond which dysfunction would be expected. Furthermore, accurate assessment allows improvements or deterioration of strength related to treatment or disease progression to be tracked and documented. Future research should address the specific impact of orofacial weakness on speech and swallowing functions, and the effectiveness of prescribing exercises to improve orofacial strength for functional gains.

Acknowledgment

This research was supported in part by NIDCD Grant R03 DC06096 and the Department of Clinical Investigation at Walter Reed Army Medical Center. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army, Department of Defense, or the U.S. Government.

REFERENCES

- Barlow, S. M., & Muller, E. M. (1991). The relation between interangle span and in vivo resultant force in the perioral musculature. *Journal of Speech and Hearing Research*, 34, 252-259.
- Barlow, S. M., & Rath, E. M. (1985). Maximum voluntary closing forces in the upper and lower lips of humans. *Journal of Speech and Hearing Research*, 28, 373-376.
- Bu Sha, B. F., England, S. J., Parisi, R. A., & Strobel, R. J. (2000). Force production of the genioglossus as a function of muscle length in normal humans. *Journal of Applied Physiology*, 88, 1678-1684.
- Clark, H. M., Henson, P.A., Barber, W. D., Stierwalt, J. A. S., & Sherrill, M. (2003). Relationships among subjective and objective measures of tongue strength and oral phase swallowing impairments. *American Journal of Speech-Language Pathology*, 12, 40-50.
- Clark, H. M., O'Brien, K., Calleja, A., & Newcomb, S. (submitted). Effects of directional exercise on lingual strength.
- Clark, H. M., Solomon, N. P., O'Brien, K., Calleja, A., & Newcomb, S. (2008, March). Lingual and buccal strength: Innovations in clinical assessment. Poster presented at the Biennial Conference on Motor Speech, Monterey, CA.
- Hayashi, R., Tsuga, K., Hosokawa, R., Yoshida, M., Sato, Y., Akagawa, Y. (2002). A novel handy probe for tongue pressure measurement. *International Journal of Prosthodontics*, 15, 385-388.
- Luschei, E. (2008). Lateral tongue strength. Retrieved April 4, 2008, from http://www.iopi.infoaateral_tongue_strength.htm
- O'Day, C, Frank, E. M., Montgomery, A., Nichols, M., & McDade, H. (2005). Repeated tongue and hand strength measurements in normal adults and individuals with Parkinson's disease. *International Journal of Orofacial Myology*, 31, 15-25.
- Robin, D. A, Somodi, L. B., & Luschei, E. S. (1991). Measurement of tongue strength and endurance in normal and articulation disordered subjects. In C. Moore, K. M. Yorkston, & D. R. Beukelman (Eds.), *Dysarthria and apraxia of speech: Perspectives on management* (pp. 173-184). Baltimore: Paul H. Brookes.
- Thompson, E. O, Murdoch, B. E., & Stokes, P. D. (1995). Tongue function in subjects with upper motor neuron type dysarthria following cerebrovascular accident. *Journal of Medical Speech-Language Pathology*, 3, 27-40.
- Wood, L. M., Hughes, J., Hayes, K. O., & Wolfe, D. L. (1992). Reliability of labial closure force measurements in normal subjects and patients with CNS disorders. *Journal of Speech and Hearing Research*, 35, 252-258.